

Can we
measure
ability ?

54
39

699

A. F. WATTS

3L
NET

~~694~~

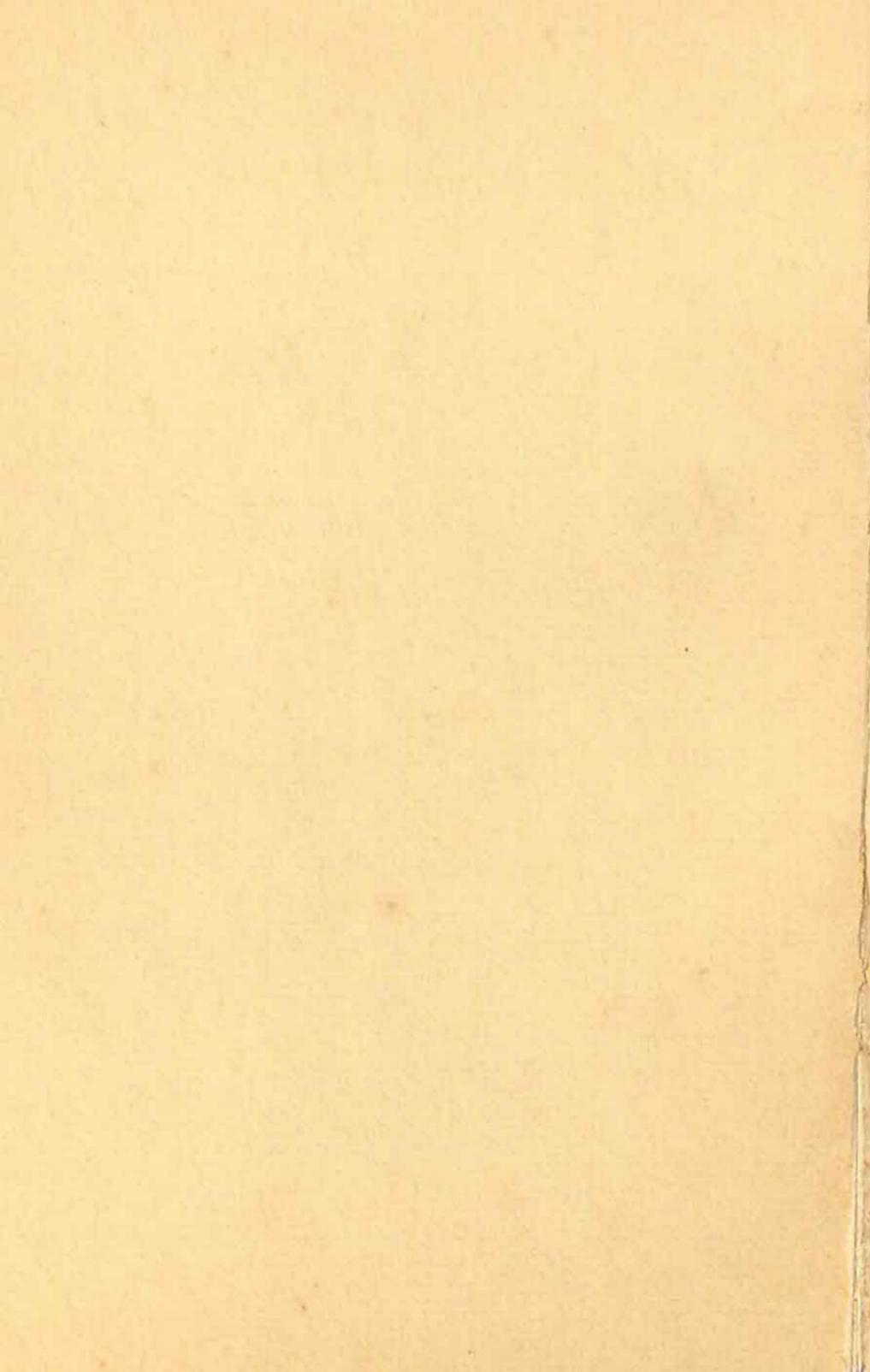
~~4199~~

~~694~~

224

~~P.S. 39~~





CAN WE MEASURE ABILITY?



CAN WE MEASURE ABILITY?

A. F. WATTS

M.A., D.Lit.

*Senior Research Officer, National
Foundation for Educational
Research in England
and Wales*



UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4

Copyright 1953
by A. F. WATTS

15.9.93
7392

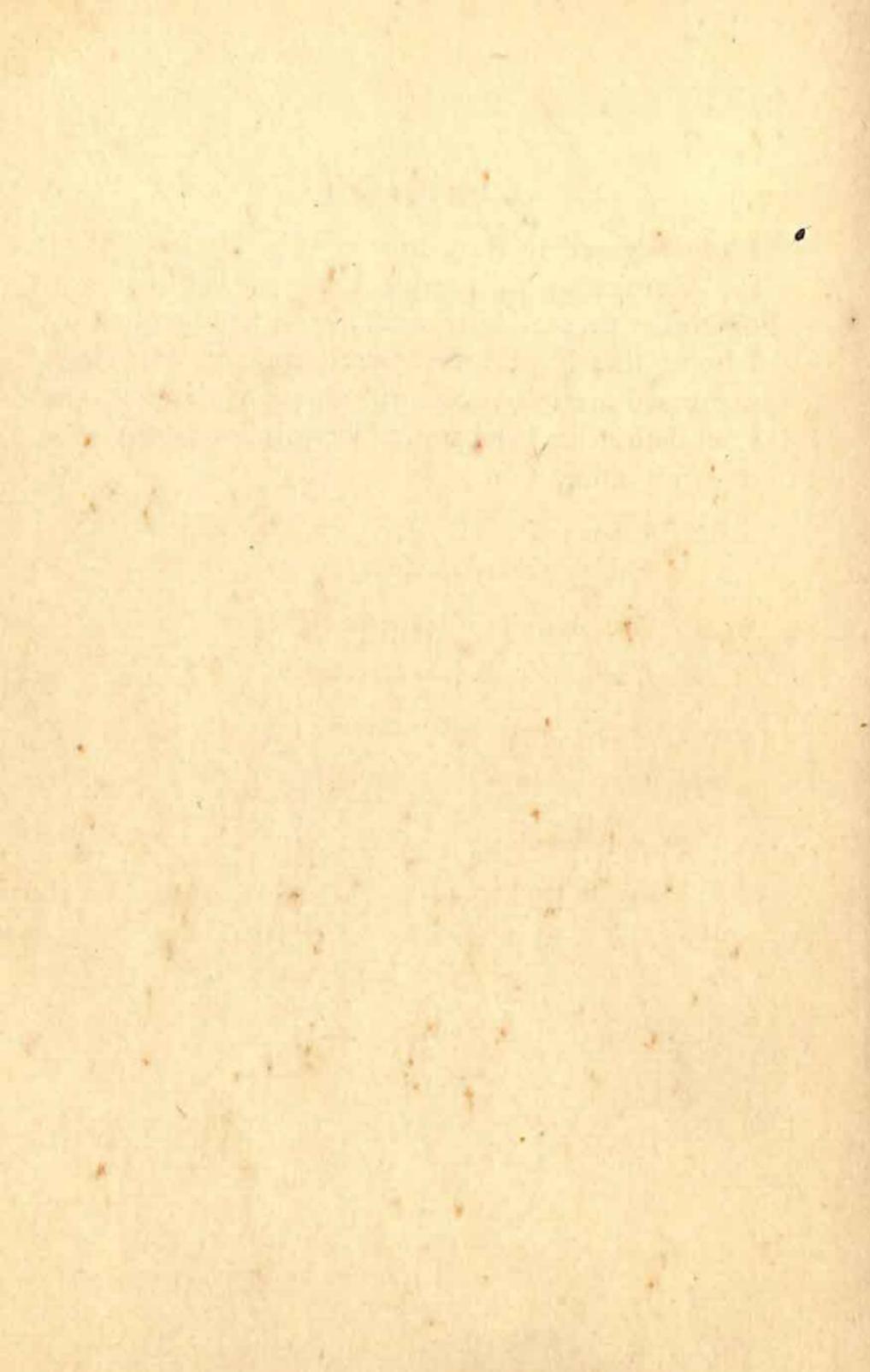
Printed & Bound in England for the UNIVERSITY OF LONDON PRESS, LTD.,
by HAZELL, WATSON & VINETY, LTD., Aylesbury and London

PREFACE

I am indebted to the editor of *The Teacher's World* for permission to reprint the material that was originally prepared for serial use. It will be obvious, I hope, that I speak for myself and that the views expressed are not necessarily those of the National Foundation for Educational Research, whose senior research officer I am.

March, 1953

A. F. W.



CONTENTS

Preface	5
1 What do Intelligence Tests Measure?	9
2 How Intelligence Tests are Made	18
3 What about Coaching?	24
4 Is Coaching Fair?	30
5 Intelligence Test Scores	36
6 English and Arithmetic Tests	41
7 The Assessment of Attainment in English	47
8 The Problem of the Late Developer	52
9 Assessments made by Teachers	57
10 The Importance of the Interview	63
11 Special Abilities	69
12 Looking to the Future	75



Chapter 1

WHAT DO INTELLIGENCE TESTS MEASURE?

This book was written as the result of an interview between the author and two others, a parent and a teacher. The interview is summarised in this Chapter.

Parent: Are you able to assure me, Dr. Watts, that the sort of examination which my child will be taking very shortly will reveal his abilities in their true form?

Teacher: Perhaps I ought to explain that my friend has been reading some of the newspaper articles on this topic which have been appearing lately.

Parent: Yes, and I think you will agree that they aren't particularly encouraging. Even your own recent letter in *The Times*, Dr. Watts, was not completely free from doubts and hesitations.

Dr. Watts: Yes, I know. But remember what was said about a famous debate on a government report some years ago—that there were three classes of people concerned in it—those for; those against; and those who had read the report. This would very well apply to a lot of the arguments you are talking about. Very few of the people who rush into print to condemn intelligence tests, for in-

stance, appear to have much understanding of how they are made and applied. I have heard their views described as austerity views; that is, they are held by those who forgo the luxury of factual or logical support.

Teacher: As far as that goes there are lots of people—like myself—who think “objective” tests may be useful, but who still don’t know enough about them to be able to counter a lot of this mischievous propaganda which is bandied about nowadays. I’ll tell you one of my main difficulties in trying to understand this business: how is it possible to measure something which you can’t even define? After all I’m not exaggerating too much when I say that there are nearly as many definitions of intelligence as there are psychologists?

Dr. Watts: Rather more, I should imagine. You are forgetting all the journalists.

Teacher: All right then. If what you are trying to measure is so vague and imperfectly understood, how can you set about estimating its quality or strength?

Dr. Watts: Let me ask you a question before I answer that one. Which came first—a precise, generally accepted definition of electricity, or a technique for

measuring the strength of an electric current?

Teacher: Yes, I think I see what you mean.

Dr. Watts: I think the two problems are very similar. Nobody knows precisely what intelligence is, or if it is such a simple thing as is implied by the tests, but that doesn't prevent us from trying to measure its effects in certain directions. Well, we can make that problem the first item on the agenda, so to speak. I think that, in order to answer our friend here satisfactorily, what we need to do is to split up his general question into more specific ones. Since so-called tests of intelligence form an important part of the examination which his son will have to face, I think that this question —what do intelligence tests measure? —might be a good one to start with.

Parent: I should certainly be glad to hear your views on that. In my day examinations in the three R's were considered good enough. Do intelligence tests measure something different?

Dr. Watts: Not altogether. The abilities employed in working an intelligence test enter very largely into tests of Arithmetic and English.

Parent: Then why bother? Why not just use

Arithmetic and English tests as we used to?

Dr. Watts: Well, let's consider some of the possible implications of doing as you suggest. Suppose that amongst the candidates for the examination there were two boys who were equally bright, two boys that is who might be expected, if given identical treatment, to progress at exactly the same rate. But suppose that one of them goes to a good school where there are small classes, efficient and enthusiastic teachers; moreover he enjoys good health and rarely misses school, and his parents give him every encouragement at home. Now suppose that the other boy is less fortunate; he goes to a school where there is serious overcrowding, where the teachers have to cope as best they can with classes of fifty or sixty children; suppose, too, that he meets with an accident or illness which keeps him from going to school for long periods, and that his parents take little interest in his progress. You will agree that it is reasonable to suppose that these two boys are likely to put up markedly different performances in tests of English and Arithmetic?

Parent: Yes, there's no doubt about that.

Dr. Watts: Now suppose we could design a technique for assessing the abilities of those two boys that would go beneath the surface, as it were, one which would ignore the effects of their training in the basic school subjects and which would measure something of their native ability. Such a test would, you will agree, assist us in trying to estimate their chances of success in, say, a grammar school.

Parent: Certainly. Is that what intelligence tests can do?

Dr. Watts: That is what the perfect intelligence test would do. The tests we use go only part of the way in that direction.

Teacher: That's all very well. I fully agree that there should be some way of taking a boy's innate ability into account, but what guarantee can you give that the tests in current use bear any relation to it. It's one thing to say that innate ability ought to be measured, quite another to prove that it *is* being measured.

Dr. Watts: I agree. What can be demonstrated, however, is that the tests we are discussing provide the best indication of the basic underlying ability which enters into every branch of school work.

Teacher: How can that be demonstrated?

Dr. Watts: Am I allowed to introduce mathematics?

Parent: No, thank you very much!

Dr. Watts: I thought not, but there is a term used extensively by statisticians which I should like to introduce and explain. It describes a very simple notion but it helps enormously in understanding the problem we are considering.

Teacher: All right, but don't expect too much, and no mathematical formulæ.

Dr. Watts: I can promise that. The concept I should like to introduce into the discussion is that of *correlation*. Any two variable quantities which are dependent on each other or are related in any way are said to be correlated. For instance, there is an obvious relationship between the level of mercury in that thermometer on the wall there and the temperature of this room. The two vary in the same direction—as the temperature rises or falls so does the mercury level. These two therefore are said to be *positively* correlated. Now consider another example. If one pours milk from a bottle into a jug there is clearly a relationship between the amount of milk in the two vessels. But this time,

as one amount grows the other becomes less—the more the jug contains, the less there is left in the bottle. Such a correlation is labelled *negative*. Many pairs of variables are encountered of course which show no discernible relationship whatsoever. If one took a group of people and measured the size of their feet and the size of their bank balances it is unlikely that any correlation would be found. Is all that reasonably clear?

Parent: Yes, I think I get the idea, and I'll take your word for it that the amount of this correlation can be measured, but how does this tie up with proving that intelligence tests really measure this basic underlying ability I think you called it?

Dr. Watts: Well, it was shown a long time ago that all tests and examinations of school work in English, History, Geography, Science—any subject you like—are *positively* correlated. Now it has been argued from this, and I think rightly, that what this fact implies is that all such tests must have something in common and that this common ingredient is what I earlier called a basic underlying ability which enters into all the work a child does at school. Furthermore—and since

you don't want any mathematics you will have to accept my word for this—it is possible to determine for any test just how much of this common ingredient it has got in its recipe, so to speak.

Teacher: And intelligence tests have got a lot, I suppose?

Dr. Watts: Exactly. They are saturated with it to a greater degree than any other form of test or examination yet devised. Not all intelligence tests are equally saturated; that is one reason why they give different measures.

Parent: Thank you, Dr. Watts. I have a fairly clear idea now of why it is desirable to introduce intelligence tests into the selection procedure, and you have given me some insight into what it is that the tests are trying to gauge. But there are a few other questions I should like to put to you. *How* do those who make the tests set about their task of designing an instrument to measure this general ability? And do all children respond to this form of assessment? What about "examination nerves" and that sort of thing? Suppose a child does badly in the tests but does consistently well at school—what then?

Teacher: Yes, and what about coaching? How far is it possible to improve a child's score by preparing him for the test? And don't children tend to vary in their ability from day to day?

Parent: And what about children who are late developers?

Dr. Watts: Well, that's about six questions—more than enough for another session next week, and I'm sure that by then others will have occurred to you.

Chapter 2

HOW INTELLIGENCE TESTS ARE MADE

In the conversation which was reported in the previous chapter I explained that the purpose of "intelligence" testing was to secure an assessment of the general underlying ability that can be shown to be exercised in all scholastic activities. A number of questions were put to me, and I shall do my best to answer them all. The one that I am going to try to answer in this chapter was: "How do those who make the tests set about their task of designing an instrument to measure this general ability?" The questioner went on to say—and this observation is commonly made—that some of the items in intelligence tests which he had seen seemed to be far too difficult for eleven-year-old children. In answering this question, I hope to be able to convince those who offer this criticism that it is based on a misunderstanding of the methods employed in test construction.

The first step in making an intelligence test is to devise a far larger number of items than are going to be used in the final version. It is common practice, for instance, to collect together at least four hundred items when preparing a test which in its completed form is designed to contain only one hundred questions. The fact that three-quarters of the items are eventually to be discarded does not imply that little care need be taken in making the initial choice. On the contrary there are a number of principles

governing the selection of the items for the preliminary draft which can be ignored only at the risk of eventually producing a test that fails to measure up to the required standards. First, the items must be of differing levels of difficulty and yet must be appropriate to the range of ability to be found in the age group for which they are intended; for clearly a question which every child fails to answer correctly or one which is easy enough for every child to get right cannot be included in the final version, since it fails to effect any discrimination between the children working the test.

Success in selecting items which satisfy this first criterion depends upon a detailed knowledge of the capacities of children at all levels of intelligence. Secondly, the questions chosen must be of the kinds that investigation and experience have shown to be most capable of revealing the ability in which we are interested. It is quite possible to produce items which satisfy the first of our conditions in that some, but not all of the children in the age group concerned, are able to answer them correctly but which nevertheless fail to yield a measure of the general ability that we are attempting to gauge. For example, a minority of eleven-year-old children are able to deal successfully with both the following items:

(a) Fill in the blank space in the following number series:

35; 27; 20; 14; —; 5.

(b) Fill in the blank space in the following list of successive Kings and Queens: Henry VIII; Edward VI; Mary I; Elizabeth I; —; Charles I.

It is clear, however, that although those who answer (b) correctly may include a high proportion of the brightest children, chance plays a prominent part in determining whether or not a child is able to supply the required information (i.e. whether there have been lessons or not in the history of the Stuart Period). It is not difficult, therefore, to accept the statement that success in answering this kind of question is not wholly dependent on a child's level of general ability. Questions like (a) on the other hand have been found to provide measures which are highly satisfactory indications of it. The differences between the capacity of items to yield undistorted measures of general ability are not always so apparent as they are in this example, and those who construct tests have often to turn to experimental evidence as a guide to choosing types of item that most successfully serve their purpose.

Even though the test constructor may feel reasonably satisfied that he has selected some four hundred items which satisfy the conditions imposed above, he declines to accept the responsibility for making the final selection. *This is left to the children themselves.* A sample of children is carefully selected so as to be as similar as possible in age, experience, and range of ability to those for whom the test is

intended, and these children work through the four hundred items, sufficient time being allowed for nearly every child to attempt every question. The way each item behaves in this try-out is then closely examined. First, the level of difficulty is assessed; many test constructors reject items which are answered correctly by less than about 15 per cent. or more than 85 per cent. of the children. Secondly, each item's capacity to discriminate between children of different levels of intelligence is carefully scrutinised. A mathematical technique has been evolved whereby this capacity can be expressed in quantitative terms. Therefore at the end of the try-out the hundred most suitable items can be selected with a considerable degree of confidence.

The hundred-item test is usually given to another sample of children along with a test which has already been in use and has proved its worth. Any differences between the performances of the children in the two tests are then examined and further revision carried out if necessary.

The test, now in the form in which it is to be used, is next applied to all the children in the appropriate age range in some chosen area, and from the results a conversion table is prepared which enables those who use the test subsequently to translate the raw scores obtained from it into I.Q.s or into some other comparable measure that makes appropriate allowances for the differing ages of the children who work the tests. By this means it is possible to compare the score of, say, a child of

eleven years two months with the scores achieved by hundreds of other children of exactly the same age. This enables us, we think, to make our comparisons just.

Finally, the "reliability" of the test is calculated. This is an index of the degree of consistency found when the test is repeated. It is clear that a test, however satisfactory in every other respect, would be of very little practical value unless we could feel confident that it would produce approximately the same order of merit if given to the same group of children on two separate occasions. The reliability of a test is calculated by a process too complicated to deal with here and is expressed on a scale ranging from 0 to 1. I shall have more to say about this feature of tests in other connections later, but it will perhaps be one indication of the rigorous standards nowadays applied to intelligence tests to state that no reputable organisation would submit to a local education authority for its use any test whose coefficient of reliability fell below 0.95, perfect reliability being represented by 1.0.

In answering this question, the first of a series of questions put to me concerning intelligence tests, I have tried to give a sufficiently detailed outline of the methods employed and the precautions observed in constructing these tests to prove that some, at any rate, of the arguments brought against them betray an incomplete acquaintance with the facts. It should now, I hope, be easy enough to dispose, for instance, of the criticism implied in the observation quoted in

the first paragraph. We often encounter the complaint from parents, and sometimes from teachers, that particular items in a test are far too difficult for eleven-year-old children. Clearly the people to blame, if this is so, are those eleven-year-old children who were so perverse as to be able to deal with these items successfully and who were therefore responsible for the decision to include them in the test.

One of the other questions put to me recently was: "What about coaching? How far is it possible to improve a child's score by preparing him for the test?" A lot of opinions and a few facts have recently been published on this issue, and I shall try in the next chapter to assess the seriousness of the effects of coaching on test scores and to discuss some of the measures which have been put forward for dealing with this problem.

Chapter 3

WHAT ABOUT COACHING?

The second question I was asked, you will remember, was about the unfairness of coaching some children only and not others. This question deserves careful consideration, because most local education authorities nowadays use intelligence tests as part of their procedure for allocating primary school leavers to different types of secondary school. It is therefore important to try to gauge the effects of coaching on the performance of children in these tests. Our difficulty is that many of the answers given by those who have investigated the problem appear to be inconsistent and inconclusive.

One reason for the apparent confusion is that coaching is an umbrella term covering a variety of processes. Suppose, for example, that a particular school contains two classes of children, which are comparable in every way, and that both these classes are coached for the same length of time. Should they both show the same effects? Imagine that one class is coached by a visiting stranger, a university don of forbidding and unprepossessing appearance and mode of address, a man who approaches the task in a detached, scientifically neutral manner, who is inexperienced in dealing with young children and uses a vocabulary which is largely beyond their understanding; who is, moreover, incapable of concealing his impatience with those unable to make rapid progress. The other class, let us suppose, is

coached by a teacher who knows the individual children intimately; she is popular with them, tactful and sympathetic in her handling of those who progress more slowly than the rest; who knows from experience which problems the children are likely to find novel and difficult, and has no difficulty in making herself understood by them. Few people would be surprised if the results showed that the effects of "coaching" were found to be greater for the second class than for the first. In short, the effects of coaching must vary with the teacher. This point, however, has been disputed by at least one expert, but you may agree that ordinary common sense has the answer.

Another common error which characterises many attempts to demonstrate or estimate the effect of coaching is to assume that increased test scores can be wholly ascribed to the conscious efforts made by the teacher in trying to bring about an improvement. I once met a teacher who claimed to have proved that coaching was effective because he had given an intelligence test to his class, with an average standardised score of 102 as a result and had then, after three or four lessons devoted to coaching, given a second test which yielded an average of 107. An obvious improvement. The nature of the logical fallacy involved in this kind of reasoning can best be understood by considering a further example. Dr. X has a theory about the cause of varicose veins. He believes that this complaint is due to wearing shoes which are too tight. In order to test

his theory he takes careful note of the size of the feet and of the shoes of all those patients he encounters who suffer from varicose veins. After a period of years, during which he has met a large number of such cases, he examines the results and finds that no fewer than 90 per cent. of his patients are wearing shoes which appear to be at least one size too small. He feels satisfied therefore that his theory has proved to be correct. Subsequently, however, his investigation is repeated by Dr. Y, who examines also an equal number of people similar in age and in every other respect except that they do not suffer from varicose veins. Dr. Y discovers that of this group, too, 90 per cent. are wearing shoes at least one size too small. What of Dr. X's theory now?

Clearly in all investigations of this kind, what research workers call a "control" group is essential if we are to make correct inferences about the changes that have taken place in the "experimental" group in which we are interested. If the teacher, to whom I referred earlier, had tested another similar class on two separate occasions he might well have found an equivalent increase in their average score without the intervention of coaching. This does actually occur, and frequently. Even if a control group is used, care is needed to ensure that its treatment is comparable with that given to the experimental group. For instance, in a coaching experiment the children in the coached group are almost inevitably encouraged to give of their best and are made to feel

that a special interest is being taken in their progress; if steps are not taken to induce similar incentives in the control group, the assumption that any differences between the performances of the two groups can be ascribed to coaching alone is considerably weakened.

I have elaborated these points at some length in order to stress the importance of examining the evidence very carefully before accepting as valid some of the claims made on behalf of the effectiveness of coaching and also in order to account for some of the differences between estimates of the size of the gains which might be expected after coaching has been given to a group of children.

It was, of course, about the size of the expected gain that my questioner was seeking information when he asked in our conversation some weeks ago "What about coaching?" and I would be doing less than justice to his question if in my answer I avoided making some assessment of the possible extent of coaching effects. It is clear, however, from all that I have said above that no figure can be quoted that could apply to all situations. The size of the increase in scores brought about by coaching depends both upon the kind of person who does the coaching and upon the particular children coached. We can quote our assessments only as averages or as upper and lower limits.

One such estimate made recently by Professor P. E. Vernon prompted a prolonged and lively correspondence in the columns of the *Times Educational*

Supplement. He suggested that coaching is likely to produce a gain of 14 points of standardised score, rising to as much as 18 points for a group of bright and unsophisticated pupils and falling to 12 points where the coached group is comparatively dull and is familiar with intelligence tests. My own experience of coaching experiments and my acquaintance with the reports of the investigations of others lead me to regard this estimate as exaggerated and I feel that in arriving at it Prof. Vernon has attached too much significance to the large gains which one of his students, Mr. D. T. Navathe, reported from experiments involving children in preparatory schools. Such children cannot be regarded as representative of the general school population and only guarded inferences should therefore be drawn from results in these circumstances.

A better indication of the size of improvement which coaching is likely to produce is provided, I feel, by an investigation conducted in Belfast some time ago by one of my colleagues, Mr. Alfred Yates. Yates was careful to select a sample of children comparable in range of ability with the group of children who normally enter for secondary school examinations at the age of eleven, and which included the correct proportion of "borderline" cases. The test used, the Otis Advanced, is notoriously more "coachable" than, for instance, the tests in the Moray House series. In spite of this, a programme of intensive coaching lasting over seven hours in all, using material closely parallel to that found in the test

proper, produced an average gain of only 9 points of standardised score from children who were as lacking in experience of tests as any likely to be found in schools nowadays. The children in the control group, it must be noted, who did two tests only and had no coaching between them, secured a gain of over 5 points. The net benefit to the coached group, therefore, was less than 4 points. Moreover, as Yates himself has admitted, it is possible that even this gap might have been reduced if the control group had been allowed further opportunities for practice. Mr. Stephen Wiseman, of Manchester, and Dr. J. J. B. Dempster, of Southampton, have since conducted carefully devised investigations which tell much the same story.

The National Foundation for Educational Research has published an account of some recent experiments which, I think, confirm the view that the 9 points of gain found by Yates, Wiseman and Dempster probably represents the upper limit of the *average* increase which can be expected to result from intensive coaching.

Let us, then, accept it as a likely thing that coaching can bring about an increase in a child's score in an intelligence test. The next crucial point to consider is the effect on examinations for the selection of eleven-year-old children for secondary school courses that can be produced by a coaching gain of the order of, say, 5 to 9 points. Assuming that some of the entrants are coached and others are not. I propose to examine this problem in the next chapter.

Chapter 4

IS COACHING FAIR?

In the previous chapter, after discussing some of the recent investigations into the effects of systematic coaching on performance in intelligence tests, I ventured to conclude that an increase of 9 points of standardised score was probably the maximum *average* gain that coaching a group of the average run of children could be expected to bring about. What is much more difficult to decide is how far a gain of this size can disturb an order of merit in a grammar school selection examination for which some children have been coached and others have not.

In a recently published pamphlet* Mr. W. G. Emmett of Moray House has attempted to answer this question. He has calculated the number of candidates out of every thousand entering for an examination who would obtain grammar school places as a result of coaching but who would not otherwise do so. He offers a number of different estimates varying in relation to the assumptions that can be made about the number of grammar school places available, the proportion of coached children and the size of the gain attributable to coaching.

For example, in an area in which 15 per cent. of the entrants are admitted to grammar schools, Emmett estimates that "under the worst possible

* "The use of intelligence tests in the 11+ transfer examination" (University of London Press, Ltd.).

circumstances," that is, assuming that half the children are coached and half are not and that the average gain produced by coaching could be as much as 14 points of standardised score, "only 13·5 children out of a thousand are favoured." If, however, as he thinks likely, the proportion of children coached is, in most areas, less than 10 per cent., and if the average gain brought about by coaching is 10 points or less, he is prepared to conclude that "the number unduly favoured is almost negligible."

Although Emmett's arguments are persuasive and firmly based, there is, I feel, a trace of unjustifiable complacency about his conclusions, and I am unable to agree with them unreservedly. Clearly, whether or not one can accept as "negligible" the errors of selection introduced by differential coaching or by any other agency must depend on the point of view one is able to adopt.

The administrator, who realises that his methods can never be perfect but who takes pains to ensure that injustice is reduced to a minimum, may well feel satisfied if he can be assured that only a few children in every thousand are being wrongly allocated; on the other hand, the teachers of these misplaced children may be more inclined to criticise the procedure; whereas the parents of uncoached and unsuccessful borderline children will certainly refuse to regard the errors as unimportant. Moreover, although some have undoubtedly erred in the opposite direction by exaggerating the effects that differential coaching can have on an order of merit,

it seems to me that Emmett's estimates tend to mask the seriousness of the disturbance.

If (returning to his example) 13 children in every thousand can in certain circumstances rise above the pass mark as a result of coaching, there must be a corresponding number who are displaced by the lack of it. Furthermore if the number of misplacements were related not to the whole age group but to that section of it which is to be found clustered around the borderline between success and failure, the size of the proportion affected would be more alarming. All this, however, assumes that selection for grammar schools is made on the basis of test scores alone. This is not so in the vast majority of areas in our country. Other things, including the recommendations of teachers, are also, as a rule, taken into account.

It is, however, less important, to my mind, to try to determine the precise number of children who might be unfairly treated by any system of selection than to face the seemingly incontestable fact that the possibility of error exists and to seek appropriate and effective remedies. If we were dealing with, say, the percentage of faulty articles produced by some technique of manufacture it might be possible to describe some very small proportion as "negligible." In considering what is a highly important decision affecting the education and, possibly, the vocational destinies of a number of children such an epithet is inadmissible. If the methods of selection in common use, whilst soundly conceived and producing, for

the most part, satisfactory results, do allow of a margin of inaccuracy, we are obliged to examine the possibilities of both minimising the errors that occur at the time of the examination and of considering the steps necessary after the event for putting right the mistakes which are found to have been made.

A ban on coaching, sustained by "gentlemen's agreements" between teachers and local education authorities, and by propaganda aimed at persuading parents that attempts to secure a grammar school place for a child who is incapable of the sustained effort required for success in an academic course are not in the best interests of the child, has been advocated as one solution. Whilst sympathising with the good intentions of those who make this proposal, I fear that it is scarcely practicable, particularly in those areas where there is severe competition for a relatively small number of available grammar school places.

I have less sympathy for the alternative suggestion that coaching for all should be introduced as an official policy. Apart altogether from the deplorable effect which such a measure could have on the life of many primary schools, the claim that it would serve the interests of fairness is based on a false logic. There is enough available evidence to satisfy me that teachers differ markedly in their ability to bring about increased test scores by coaching, and so universal coaching would reintroduce into the selection procedure those very injustices which the use of intelligence tests was intended to mitigate.

Again, since individual children differ in their susceptibility to the effects of coaching, irrespective of their level of initial ability, a policy of coaching for all would confer advantages on those children who are capable of making relatively large and rapid gains during a brief spell of preparation. There is no evidence to show that such children have more capacity than others for the sustained and successful prosecution of an academic course of secondary education.

There would, fortunately, appear to be growing evidence in support of the claim that unassisted practice can be just as effective as an equivalent amount of coaching in bringing about an improvement in the scores of a group. This, then, could well be the happiest solution to the problem of how best to minimise the errors introduced into the selection procedure by the fact that some coaching will probably go on in spite of attempts to discredit it. Its effects could be offset by the provision of the opportunity in each school for working three or four practice tests, and the arrangement for this would be far less disruptive than the introduction of coaching into the curriculum. Moreover, the differences between schools which affect the gains brought about by coaching could scarcely influence the results of practice in the same way.

I do not suggest for a moment that this or any other remedy will enable us to differentiate between the abilities and characteristics of children without some error, but a consideration of the steps which

can and should be taken after the event to repair the deficiencies of selection must be postponed. In the next chapter I propose to deal with some more of the questions that have been put to me concerning the effects on the attempted assessment of a child's capacities of factors, such as "examination nerves" and the tendency of children's effective ability to vary from day to day.



Chapter 5

INTELLIGENCE TEST SCORES

I stated at the outset that a child's score in an intelligence test cannot be regarded as a precise measure of his native ability. If the test has been efficiently constructed and properly standardised it provides us with a means—as yet the best available means—of making a judgment about his capacities, but it would be foolish to pretend that the size of the score is determined solely by the powers we are seeking to assess. I have referred already to two factors that are capable of influencing the final result: (a) even the most carefully designed tests fall short of complete reliability and so admit of random errors; (b) coaching for and practice in the working of tests, if enjoyed by some and not by others, can falsify our attempts to make just comparisons between the "real" abilities of individuals in a group of children who have not all been treated alike as regards coaching.

Recent controversies have revolved around these two factors so exclusively as to give the impression that no other sources of error exist. I cannot believe, however, that even if a perfectly reliable test could be devised, and one, moreover, that was proof against the disturbing influences of coaching and practice, we could, without reservation, regard the results obtained from it as being faultless indicators of a person's true capacities. On the contrary, I am satisfied that there are two, possibly related, sources of error

which have hitherto received scant attention. The first is the personal variability which is characteristic, most people would agree, of every aspect of human endeavour. We accept, without undue surprise, the fact that a first-class batsman might score a century and a "duck" on successive days, even when there is no apparent variation in the state of the pitch or the standard of the bowling on the two occasions. Although an ability may be relatively constant as far as potential achievement is concerned, the fact that its day-to-day effectiveness varies is a matter of common observation. We say, for instance, of a child with whose progress we are intimately acquainted that we know what he is "capable of"; there are not infrequently, however, occasions when, after examining a particular specimen of his work, we feel constrained to tell him that he "can do better." That this variability affects the scores obtained in intelligence tests is supported by evidence provided by some of my own experiments. An account * has been given of one investigation in which ten examinations, each consisting of an intelligence test and tests of attainment in Arithmetic and English, were given weekly to several groups of children. The fluctuations of the scores of some of the children in all three tests were far greater than could be accounted for by the fact that the tests used were not perfectly reliable. It seems clear, then, that a child can be "off form" when working an intelli-

* *Secondary School Entrance Examinations*, London (Newnes, 1952, 3s. 6d.).

gence test and, therefore, the score derived from a single administration does not provide us with a satisfactory basis for assessing his real capacities.

The second potential source of error to which I wish to refer—and this, as I said earlier, may well be related to the personal variability in performance that I have just been discussing—is to be found in the fact that the processes involved in working intelligence tests do not take place in some self-contained compartment of the mind which is insulated from the effects of prevailing mood and emotional excitation. Every experienced teacher knows that a child's normal school work varies in quality according to whether he is confident in his capacity to succeed or inhibited by a sense of inadequacy; sustained by an ordered and congenial home background or discouraged by being deprived of parental encouragement and affection; happy or depressed; persistent in the pursuit of worthwhile goals or feebly motivated and easily distractible; blessed with abundant vitality or crippled by chronic lassitude. It is scarcely feasible that intelligence tests alone should be immune from such influences. There is not a great deal of experimental evidence available that bears on this point, largely because it is difficult to devise suitable methods for the investigation of the effects of differences in personal make-up or home background on performance in intelligence tests. One can fairly easily arrange, for example, for some coaching to be given to a group of children and to be withheld from a comparable

group in order to compare their subsequent performances, but an examination of the effects of, say, undernourishment or anxiety cannot be so readily contrived. Such investigations as have been carried out, however, have served to reinforce the conclusions arrived at by the exercise of common sense. One experiment carried out by E. B. Hurlock consisted in testing three matched groups of children, one of which was praised after the first effort and encouraged to try to do even better on the next occasion, another was severely reprimanded and told that if higher scores were not achieved in the second test punishment would follow, while the third (a control group) was retested without any comment being made about the scores achieved in the first test. The results of the second test showed that both the praised and the reproved groups made significantly greater gains in score than did the control group.

Another investigation demonstrated the effect on test scores of what has been called the "level of aspiration." Let me dissociate myself in advance from any suspicion that I approve of the methods employed. Several groups were tested and were then told what scores they had obtained and what scores other groups had secured. In some cases the information given was accurate, in others the groups were told that they had done better or worse than they had actually done. Similarly the scores made by the other groups were in some cases exaggerated and in others understated. Each group was then asked to estimate what their scores would be in a

second test. The results showed clearly that the increases in score in the second test were related to these estimates of their own potentialities which in turn were shown to be dependent on the information that had been given to the groups about their own scores and those of other groups. In other words the morale of the groups concerned and the attitudes induced in them towards their own achievements and potential improvement had a significant effect on their subsequent performance.

I shall shortly be giving details of the great increases effected in children's scores by their gaining greater self-confidence in their work.

The consideration of these sources of error is not intended as an attack on intelligence tests—any alternative method of assessment is just as liable, possibly more liable, to be affected by these factors. The wholesale condemnation of intelligence tests is plain silly. What I do wish to stress is that the result of a single intelligence test can do no more than help us to classify children into a limited number of groups, and fine discriminations, particularly between children whose scores are separated by only one or two points, cannot be justified.

So far, I have dealt exclusively with intelligence tests. The methods used in assessing children's proficiency in employing the basic scholastic skills also deserve examination. In the next chapter I propose to discuss the kinds of test that are commonly used nowadays to measure attainment in English and Arithmetic.

Chapter 6

ENGLISH AND ARITHMETIC TESTS

The attempt to measure attainment is not a new feature of educational practice. Few can ever have taught successfully without pausing from time to time to find out how much information their pupils have acquired and retained, how capable they are of exercising the skills to which they have been introduced, and how far they can utilise what they have learned in addressing themselves to novel problems. The traditional methods of assessment consisted in framing a number of questions based on the subject matter under review and in getting the pupils to answer these orally or in writing. The kinds of attainment test that are now in common use share these features with their predecessors but differ from them in being "standardised" and "objective."

By a standardised test we mean one—as explained at the outset—which prior to its use has been tried out extensively on children comparable in age and range of ability with those for whom it is intended. As a result we can relate the score that a particular child obtains to those achieved by other children of exactly the same age. It is important, however, to recognise that age is the only variable for which standardised tests make due allowance. They afford no compensation, for instance, to the children whose schooling has been interrupted by protracted illness, or to those who have been the victims of in-

different teaching or of an uncongenial home environment.

An objective test is one of which the questions are as far as possible unambiguous and are designed to evoke brief answers that can be easily and accurately marked. It is well known that if the same examination papers containing "essay-type" questions of the traditional kind are marked by different examiners the scores assigned to them will vary, sometimes considerably. Objective tests, on the other hand, yield an almost perfect consistency between the scores allotted by different markers.

The advantages that can be derived from the use of carefully prepared, standardised, objective tests are easily recognised by teachers, and it is not necessary for me to elaborate the arguments in favour of their use. Indeed, I am inclined to believe that there is nowadays a danger, not that teachers may overlook the merits of the newer types of attainment test, but rather that they may fail to recognise the fact that there are some occasions when it is not essential to employ them and that there are others when it is definitely undesirable to do so.

Clearly, when tests are to be used within a school or within a class in order to arrive at an order of merit on the basis of which children are to be allocated to grades, sets or "streams," it is by no means essential that standardised tests should be used, although the merits of making one's assessments as objective as possible are apparent enough. The purpose in this instance is to distinguish be-

tween the performances of the children concerned and the relation of their scores to those achieved by children of the same age in another area is irrelevant. Furthermore, there are abilities and achievements that we need to estimate for which objective tests are wholly unsuitable. There is no escape, it seems to me, from the subjectively assessed "essay-type" examination, however unreliable it may be, when we wish to estimate, say, a pupil's capacity to marshal arguments in support of some point of view or the ability to arrange ideas into new and individual patterns. These are abilities to be found only perhaps in a selected minority of older school-children, but it is necessary, nevertheless, to define the limits of the usefulness of objective tests since some of their more enthusiastic devotees appear to be prepared to introduce them at every level.

The attainment tests that form part of the examinations organised by local education authorities for assigning primary school leavers to appropriate secondary school courses clearly need to be standardised in order that due allowances may be made for the differences in age between the candidates. Whether or not the tests should be wholly objective is, perhaps, open to argument. Before entering upon this argument, however, let us consider the assumptions underlying the use of attainment tests in making comparisons between children from different schools.

Can we reasonably assume, for instance, that all primary schools follow the same curriculum in Eng-

lish and Arithmetic or devote equal proportions of their time to teaching these subjects? In some countries there is so rigid a central control exercised over the schools that an affirmative answer can be given to these questions; we, however, pride ourselves on the freedom enjoyed by the heads and staffs of our schools to provide, within generous limits, a curriculum suited to their own predilections. In such circumstances can a common test of attainment be said to be fair to all the children who are assessed by it? One way of resolving this dilemma is to base attainment tests on the minimum of knowledge and skill that all children can be supposed to have acquired by the age of eleven, no matter how unorthodox or even bizarre their educational treatment may have been. This, in fact, would appear to be the only alternative to the imposition of an "agreed" syllabus on all primary schools. It would seem to be along the former lines that we have attempted to solve the problem caused by the diversity of practices that obtain in our primary schools. Attainment tests for Arithmetic, for example, for the most part demand little more than a knowledge of the four basic rules, and problems involving, for instance, a knowledge of decimal fractions are omitted because some primary schools follow a curriculum that does not introduce them.

One might have expected that this severe restriction on the range of questions included in modern attainment tests would have led the staffs of primary schools to regard themselves as free to

fashion their own preferred programme, secure in the knowledge that the examinations awaiting their pupils are of the kind that any well taught child should be able to take "in his stride," so to speak. In some areas, however, where competition for a few coveted grammar school places is exceptionally severe, a different and most unfortunate consequence has followed upon the introduction of attainment tests of the modern type. Some teachers, anxious, no doubt, to serve what they conceive to be the best interests of their pupils, have modified the curriculum of their schools, particularly during the final year before the examination, until its contents have corresponded closely to the minimum required for successfully working the attainment tests to be encountered. Thus, ironically enough, attempts to avoid imposing on the primary schools a syllabus which some teachers would find irksome in that it hampered their efforts to provide for their children a comprehensive and richly varied educational programme, are actually leading in some areas to the adoption of a prescriptive routine of narrow specialisation.

What can be done to halt this tendency which, if it were allowed to continue, could have a deplorable effect on the well-being and educational standards of our primary school children? Clearly the attainment tests—since experience has shown that they exercise an influence on the junior school curriculum—must be modified.

Tests of attainment in Arithmetic could, I think,

without change in format or loss of objectivity, be progressively broadened little by little and by agreement in order to provide for the assessment of skills that most teachers agree can be successfully taught in primary schools, but which are not measured by the kinds of test used nowadays in most areas.

I am not satisfied that similar minor modifications would be sufficient to rectify what are, to my mind, serious imperfections in the methods used currently for the measurement of attainment in English.

THE ASSESSMENT OF ATTAINMENT IN ENGLISH

In many areas objective tests of attainment in English are nowadays used as part of the procedure for determining the most suitable courses of secondary education to be followed by children leaving their junior schools at the age of eleven. As I have indicated, I am not satisfied that these tests best serve the purposes for which they are intended. Their main defect, it seems to me, is that they provide little or no opportunity for an appraisal of a child's ability to *use* his language "naturally" either in speech or in writing.

Admittedly it is important to assess a child's capacity for linguistic analysis—his knowledge of the meaning of words and of their function and relationship in a phrase, clause or sentence. But this does not constitute what we mean when we talk of "mastery" of one's language. A knowledge of the available tools and an appreciation of their potentialities is essential to, but is not synonymous with, good craftsmanship. Imagine an objective test of attainment in woodwork. It might contain a number of searching questions about the choice of implements and materials, the merits of alternative designs and techniques, but if it made no provision for translating theory into effective practice it would surely not be accepted by teachers of handicrafts as a satisfactory means of estimating ability in any

branch of handicraft. Indeed, we know that some of those who would be able to score highly on a test of this kind would nevertheless prove to be incapable of doing a satisfactory piece of practical work.

Is this true of some children who make high scores in the kinds of objective test of English attainment now in common use? Are some of these children unable to give clear and effective expression to their thought and experience? There is a growing belief among teachers that this discrepancy is not uncommon and I recently encountered some evidence to support it. I had the opportunity of reading a large number of essays written by children who had been allocated to secondary schools on the basis of an examination that included an objective English test but did not provide for the assessment of the capacity to write consecutive English. It was perfectly clear that quite a different order of merit would have been obtained had the essays been taken into consideration, and it was evident that a fairly large proportion of those children whose attainment in English had been rated highly were nevertheless demonstrably incapable of expressing themselves clearly or even intelligibly.

Since one of the purposes of the examination of eleven-year-old children is to select those who appear likely to profit from a grammar school course, the failure to include some method of assessing the ability to write effectively is a striking omission. For this is one of the abilities that will be exercised from the outset when the child starts his grammar school

course. Not only those activities classified under the heading of "English" in the curriculum but his progress in most of the other subjects will be grievously handicapped if he is unpractised in communicating his thoughts, unable to arrange his ideas into logical patterns, halting and imprecise in "writing up," "taking notes," furnishing accounts of laboratory experiments and in all the other diverse tasks that presuppose for their successful prosecution a considerable facility in writing.

It is perhaps no less important to ensure that children allocated to grammar schools can comprehend and respond to the spoken word. Only if the much maligned interview forms part of the procedure for estimating ability is any attempt made to assess this aspect of attainment. Most teaching, nevertheless, involves discourse and conversational response from the pupil and puts a high premium on the capacity to profit from verbal exposition. In grammar schools in particular the teachers' vocabulary and patterns of speech tend to make demands on the pupils which the verbally ill-equipped or untrained often fail to satisfy.

Clearly there are minimum levels of comprehension and expression that children need to have reached if they are to undertake a grammar school course with reasonable prospects of success, and, equally clearly, objective tests of the modern kind fail to determine whether or not these levels have been reached. This deficiency may be partly due to the fact that these minimum levels are not made ex-

plicit, let alone precisely defined. Teachers in grammar schools tend to make assumptions about what the new boy *ought* to be able to understand and to achieve, often on the basis of an inadequate knowledge of what he has been taught or of the way in which he has been taught in his junior school. Similarly, primary school teachers may entertain notions about the range of abilities and attainments that should be cultivated in children which do not necessarily coincide with the demands that will be made on them at the secondary school. On the other hand the ability of thoughtful junior school teachers to predict which of their pupils are capable of beginning a good grammar school course is remarkably successful—quite as successful as the examination scores themselves. A closer liaison between the two stages of the educational system would appear to be necessary in order to define what can reasonably be expected by the secondary schools and what can satisfactorily be provided by the primary schools. Out of this closer relationship there could emerge a blue print for a revised form of English examination which would—I think this is inevitable—probably have to sacrifice some of its objectivity in the interests of securing a wider and more valid sampling of the children's abilities to understand and to employ their language. We might even have to be daring enough to ask children to write essays and to seek for some reliable means of assessing them.

I referred last week to the general effects of objective attainment tests on the curriculum of some

primary schools. On the teaching of English, what my friend, Mr. G. F. Peaker, H.M.I., has called this "backwash," has been particularly devastating. Although it is, fortunately, untrue of the majority, there are a few schools in which the concentration of effort on the ticking, crossing out and underlining required by objective tests has left the children with little time, or, it would seem, inclination, to try their hands at writing a complete sentence. The widening of the scope of the English examination would perhaps serve to release these children from their bondage and enable them to give freer rein to their imaginative and creative powers.

In speaking of minimum levels of attainment I referred only to the requirements of the grammar schools. Perhaps the secondary modern schools too should be permitted the luxury of demanding some degree of proficiency of their entrants. Would it be advisable to refer candidates for admission to the secondary stage back to the primary school for a further period if they fail to reach certain agreed standards? I am not wedded to this idea, and indeed sympathise with a number of arguments that can be brought against it, but I think that it deserves to be discussed and put it forward for that reason.

During the conversation that was reported in the first chapter, I was asked a question about "late developers." Children certainly develop at different rates, and therefore the fact that secondary school entrance examinations are held at a fixed time raises a number of problems.

Chapter 8

THE PROBLEM OF THE LATE DEVELOPER

Children are not as considerate as they might be towards their examiners. By electing to be born on different days throughout the year they make it necessary for us to introduce age allowances into our schemes for assessing their abilities at the age of eleven. There is also some evidence that their development proceeds at different rates, and it is therefore arguable that some of them must manifest spurts of improvement after they have been examined and so may upset the order of merit on which we base our predictions about their likely progress in the future.

That there are any significant differences between children's rates of development is sometimes denied. One distinguished educationist has been heard to say that, after many years' experience, he is still looking for a "late developer." He does not claim that assessments made when a child is eleven are infallible, but asserts that he has never encountered a child of, say, thirteen who is demonstrably capable of profiting from a grammar school course, but who provided no evidence of his proclivities at the age of eleven. His true ability may not, for a variety of reasons, have been reflected in the examination scores on which the child was judged, but a more searching inquiry into his school record would, it is argued, have enabled a correct forecast to have been made.

There is evidence, however, provided by a number of investigations, which strongly suggests that

mental development is subject to the same sort of variations as have been found to be characteristic of physical growth. We know, for instance, that increases in height do not occur evenly and progressively from birth to maturity. There are two periods—one during the first few years and one at the onset of adolescence—during which growth is markedly accelerated. Whilst this trend is discernible in all individuals, there are differences between the age at which the phases occur and differences between the durations of the periods of intensive growth. Averages drawn from measurements of large numbers of children at successive stages of development mask these differences, and the curve produced by plotting these averages is smoother than that obtained from the measurements of any single individual.

Is it possible to discern any similar pattern in the development of children's abilities? There are few adequate sources of evidence, since this question can only be answered by what has come to be called a "longitudinal" study. By this we mean a prolonged investigation into the progress of a group of children, and one which involves regular and frequent measurement. The Harvard Growth Study was such an investigation. In this inquiry 3,500 children were observed and assessed at short intervals over a period of seven years. The experiment yielded far more information than has as yet been fully analysed, but there are several indications that support the belief that the abilities measured by intelligence tests and tests of attainment do not develop at a uniform

rate and that the ages at which periods of rapid improvement occur, the duration of these periods, and the age at which abilities can be said to have reached their optimal level vary from person to person.

Some striking examples of the effects of different rates of development on the scores obtained in standardised tests are given by Dearborn and Rothney.* They present their findings in statistical terms with which everyone may not be familiar, and so I have selected an example from their results and have tried to restate them in a form which, I hope, will make their significance clear.

In this example seven girls have been selected who had exactly the same score in a standardised test given to them when they were sixteen years of age. Imagine that each of the seven girls belongs to a group of 100 children and that all the groups have exactly the same range of ability. The table on the opposite page shows the position that each girl occupied in her group at eight, eleven, thirteen and sixteen years of age. (Note that the orders of merit were based on standardised scores—that is, appropriate allowances were made for differences of age at the time of the testing—and adjustments were also made to correct errors due to the differences between the tests used on the four occasions.)

Consider these results in relation to the problem of selecting those children who appear to be capable of profiting from a grammar school course. In many

* *Predicting the Child's Development* (Sci-Art Publishers, Cambridge, Mass., 1941).

TABLE ONE

Girl	<i>Position in group of 100 children</i>			
	<i>At 8 years of age</i>	<i>At 11 years of age</i>	<i>At 13 years of age</i>	<i>At 16 years of age</i>
A	1	15	3	16
B	3	16	4	16
C	33	19	15	16
D	44	42	16	16
E	50	46	16	16
F	55	52	55	16
G	58	62	59	16

areas a girl, whose abilities were such that in an order of merit of a random sample of 100 children she occupied the sixteenth position, would be regarded as suitable for a grammar school, although if she fell below this position in the list she might fail to secure a place. The seven girls in this example, when assessed at sixteen years of age, were found to have reached the level of ability usually deemed necessary for success in a grammar school course. If they had been judged at the age of eleven, however, only girls A and B would have fulfilled the necessary requirements, whereas at the age of thirteen girls C, D and E would also have succeeded. These three latter girls are examples presumably of those "late developers" for whom many authorities provide opportunities of proving their worth as "over-age" candidates in the secondary school entrance examinations. The scores of girls F and G suggest that

there are, possibly, children who fail to exhibit their powers at their full strength until they reach an age beyond that statutorily fixed for leaving school.

I freely admit that the children discussed in this example are not representative of the vast majority, and it is unlikely that predictions based on a set of examination scores obtained on a particular day during their twelfth year are unjust to more than a small minority of the children concerned. But if we are satisfied that a disservice is being rendered even to a small minority we are compelled to seek some remedy. Clearly a single examination can only provide a "cross-sectional" picture. The results tell us how a child's performance compared with those of other children *at the time of the examination*. If we are to supplement the information we obtain in order to secure some indication of the stage of development reached by a child in relation to his own potential, it is necessary to examine the record of his progress during, probably, the whole period of his primary school education.

Fortunately most local authorities are showing an increasing awareness of the importance of the detailed information about a child's development that only those who have taught him can supply, and most teachers in primary schools recognise the value of frequent estimates of a child's progress. The time is not far distant when the record card, far from being regarded as an administrator's quirk, or as an extra burden foisted upon the harassed teacher, will be considered to be the *sine qua non* of educational guidance.

Chapter 9

ASSESSMENTS MADE BY TEACHERS

In my last chapter I referred to the disadvantages that might be suffered by "late developers" if their performance in an examination held on a particular day provided the only basis on which a decision could be made about their educational future. I then suggested that information, furnished by their teachers, concerning the progress made by these children during the years they had spent in the junior school afforded the best, if not the only, means of securing a more adequate estimate of the potentialities of this small minority. It is not a novel idea, of course, that regard should be paid to teachers' estimates in this connection. Many local education authorities invite junior school head teachers to submit reports on the children who are taking part in the secondary school entrance examinations, and these reports are used in various ways, particularly when "borderline" candidates are being considered.

The Norwood committee, moreover, whose report on *Curriculum and Examinations in Secondary Schools* was submitted to the President of the Board of Education in 1941, went so far as to suggest that "differentiation for types of secondary education should depend on the judgment of the teachers in the primary schools, supplemented, if desired, by intelligence and other tests." They said that they would regard "the judgment of the teachers—based

on observation of the classroom work, the general interests and certain qualities, as for example, power of sustained effort, shown by the pupils as the most important factor to be taken into consideration in the recommendation of the appropriate education for him." Provided that such a recommendation was based on a sound and reliable school record of a child's progress and development, "compiled by teachers trained to observe and to reflect upon their observations," they were prepared to regard it as "the best single means at present available of discovering special interest and aptitude and general level of intelligence."

It is, of course, one thing to say that teachers' assessments of their pupils' abilities should be taken into consideration; it is another to determine how precisely these assessments should be made by teachers and interpreted by others who have to make use of them.

The difficulties involved in the use of teachers' estimates are well recognised. Teachers are often found to disagree about the quality of a child's work and about the level of ability required for success in an academic course. Within any school, therefore, it is feasible that two teachers might place a class of children differently in an order of merit and might not be prepared to recommend the same proportion of children for places in a grammar school. When, moreover, an examination involves children from several schools and a comparison has to be made between their performances, the problem of making

proper use of teachers' estimates is complicated further.

A teacher, whose experience has been confined to schools in which the pupils fall within a limited range of innate ability with an average level below that of the general population, may well be prone to over-estimate the capabilities of the few children who appear, in undistinguished company, to be outstanding. Similarly some children capable of profiting from a grammar school course might be rated as comparative dullards by teachers who have grown accustomed to teaching classes of highly gifted pupils.

Clearly, to overcome this difficulty, some method must be sought of adjusting the assessments in a way that will render them comparable from school to school, and it must be one that will minimise the errors produced by teachers employing different techniques of evaluation.

Two broad approaches to the problem are possible: we may accept the assessments made by the teachers and then submit them to mathematical treatment, designed to adjust them to some common scale; alternatively we may seek to devise procedures whereby the assessments will be made in such a way as to avoid the necessity for subsequent modification.

The first approach has been attempted in a number of ways. *The Scaling of Teachers' Marks and Estimates*, by McIntosh, Walker and MacKay (Oliver and Boyd, 1949), furnishes a clear account of several simple and practicable methods that can be

applied either on a large scale by examiners, or within a single school, or even within a classroom.

A number of local education authorities have evolved their own techniques for adjusting the estimates supplied by teachers. The National Foundation for Educational Research has tried out, for experimental purposes, a technique that is at once simple to apply and also appears to be successful in eliminating the grosser errors introduced by differences in standards of assessment between schools. The Quota system, originally advocated by Valentine, can be included in this category of alternatives. This method, which a small number of authorities employ and appear to find satisfactory, consists in relying on the teachers for estimates of attainment in English and Arithmetic, and in using an intelligence test worked by all the children of an area in order to make appropriate allowances for the variations in levels of ability from school to school.

The second possibility—the attempt to secure a degree of comparability between the original assessments of different teachers—has not, as far as I am aware, been adequately investigated. It seems to me that there are two ways in which the attempt might be made. The excerpt from the report of the Norwood committee, which I quoted earlier, referred to “teachers trained to observe and reflect upon their observations.” Might it not be possible to arrange for one or two teachers from each junior school to be given special training in the techniques of assessment? Such training would presumably in-

volve not only familiarisation with all the available instruments of measurement, but would also include facilities for secondment to other junior schools and to secondary schools. By this means a teacher would have the opportunity of forming a judgment both concerning the general level of ability and attainment in junior schools as distinct from a knowledge of one or two schools only, and could also relate his assessments to the requirements of the various kinds of secondary schools with which he had become familiar.

Another possible method, which I consider to be worth trying, calls for co-operation between the schools in a neighbourhood. Until recently teachers in training were inspected during their school practice by H.M.I., whose reports served to assist the training colleges in their endeavour to maintain uniform standards. Could not a similar principle be adopted in recommending primary school pupils for grammar or technical school training? I suggest that the following procedure might operate successfully:

(a) Early in the New Year primary Head Teachers would prepare orders of merit of their midsummer leavers based on their now recorded observations and on intelligence and attainment tests (including a test in English "composition").

(b) The primary Head Teachers (or experienced assistants nominated by them) would form an examining body which would divide themselves into twos, and arrange for each two to visit a different school and confer with the Head Teacher about his

(or her) recommendations for the secondary education of their leavers. They would look at the exercise books and examination papers of the children and interview difficult cases.

(c) They would then draw up, in consultation with the Head Teacher, an agreed list of recommendations which would be sent forward to the education office.

A scheme, such as is outlined above, could be put into operation at once, if, in addition, intelligence tests were used in order to allot an appropriate number of places to each primary school. The Head Teacher's agreed list, suitably quantified, would take the place of the Arithmetic and English papers.

THE IMPORTANCE OF THE INTERVIEW

Few people refrain from making private assessment of the abilities of their friends, associates and acquaintances, basing their judgments on inferences drawn from a close observation of, in particular, appearance, gestures and speech. Some claim a considerable success in arriving at a true estimate on the basis of first impressions, and assert that only rarely do they find that early, intuitive, judgments require to be revised; others acknowledge that they often feel constrained to modify their original assessments after they have had opportunities for further acquaintance with those whom they have been trying to appraise. Rarely, however, does anyone confess to a complete inability to arrive, sooner or later, at a satisfactory evaluation of another person's capabilities by watching his behaviour and listening to his conversation.

It is not, therefore, surprising to find that the interview has become generally recognised and approved as a means of differentiating between the qualities and potentialities of a number of people when a selection has to be made for some particular purpose. The general public regards the use of the interview as fair and acceptable provided that it is conducted by reasonably qualified and unprejudiced assessors. In attempting to discriminate between the educational needs and capacities of those eleven-year-old children who congregate, in an examination

list, around the border line, many local education authorities arrange for them to be interviewed by a panel of experienced teachers. This arrangement appears to meet with the general approval of parents and teachers, and considerable confidence is placed in the ability of such a panel to arrive at a valid prediction.

A growing body of criticism, however, has been levelled of late against the use of the interview as a method of selection. An impressive amount of experimental evidence has been collected in support of the contention that the interview is unreliable as a measuring instrument. One investigation, for example, showed that different panels of equally experienced sales managers placed the same group of candidates, applying for posts as salesmen, in orders of merit that bore little resemblance to each other; moreover, it was found that when a group of candidates was interviewed a second time by the same panel there were marked changes in the assessments awarded. It should be pointed out, however, that the greater part of the evidence available is derived from inquiries in which adults have been wholly concerned, and that few adequate experiments have been carried out in order to ascertain the degree of reliability that can be expected of the judgments of children's qualities made by experienced teachers by means of an interview.

Even if it could be shown that such judgments were totally unreliable, it does not seem to me to justify the abandonment of interviews altogether as

some have advocated, although we should of course be compelled to examine the ways in which interviews are conducted with a view to trying to effect some improvement. For instance, we know that individual, subjective assessments of essays and of "essay-type" questions in an examination are usually found to be unreliable, but we also know that by pooling the estimates of three or four judges a much higher degree of reliability can be achieved. This principle should be applicable to the individual judgments made by the members of an interviewing panel, and perhaps some of the disappointing results of investigations into this problem can be ascribed to the fact that satisfactory pooling arrangements were not secured. If, for example, the members openly discuss each candidate before attempting to arrive at an agreed verdict, it is quite possible for one strong-willed or opinionated individual to influence the judgment of the rest of the panel and for them to agree to an assessment markedly different from that which would have resulted from the averaging of each person's unmodified rating.

If, however, it is admitted that, whatever refinements of technique may be introduced, it is unlikely that the reliability of the interview can be made comparable with that achieved by objective tests, there is, nevertheless, still a case to be made for its inclusion in any scheme of assessment.

One possible justification is that, by means of an interview, attempts may be made to estimate certain capabilities that are not exercised by objective

tests or by written examinations. I have already referred to the fact that no provision is made in the usual form of examination for the assessment of a child's powers of oral expression or of his capacity to understand the spoken word; qualities of personality and character are measured only indirectly and certainly not adequately by written examinations; effective intelligence—that is the ability to apply intelligence to the solution of real problems—can only be inferred from the results of conventional paper-and-pencil tests. I fully realise that the attempt to form a reliable estimate by means of an interview of any of these qualities is extremely difficult, but the possibilities deserve to be investigated. I would suggest, however, that an interview board should become thoroughly aware of the pitfalls about them. A Junior Head Teacher told me recently that he had said at a meeting that he wondered how a secondary Head could differentiate in an interview in ten minutes between two children whom he could not himself place in order of priority for a grammar school place after knowing them for four years. The reply was, "I only wish it were possible for you to be present and see how we do it." Haven't we all, alas, met such wiseacres?

So far I have been considering the interview as a brief encounter between a group of assessors and a candidate.

If, however, a wider conception were to be formed of the duties of an interviewing panel, the value of the contribution, which it could make to

the appraisal of children's educational needs and potentialities, seems to me to be substantial. I envisage a panel of experienced teachers who have before them, for each child concerned, a full array of scores derived from objective tests, together with information supplied by his teachers concerning his school work (with examples of it), the trend of his progress through the junior school, and some estimate of his interests, temperament, personality and character.

The panel could interview a child not primarily in order to assess his abilities, since the evidence already in their possession should, in most instances, enable them to do that, but rather in order to be able to organise the information they possess into a better integrated and more intelligible whole. The actual interview might well be confined to securing factual information about the child's interests and experiences; these might serve, for instance, to explain apparent discrepancies between scores or ratings that would otherwise prove difficult to reconcile. And might it not, in certain cases, be useful and desirable to interview a child's teacher or even his parents in order to supplement the knowledge that the panel has gleaned about the child from other sources? Only the other day I came across an example of the need for such an approach. I was looking through the scores obtained by a group of children in a series of four intelligence tests taken at intervals during the last year of their junior school course. The scores for one girl were: 105; 106; 123;

125. The interval between the second and third tests was the six months *following* the conclusion of the local authority's secondary school entrance examinations and no coaching or practice had taken place during this period.

Without further information it would be difficult to arrive at any satisfactory explanation of this unusual trend. The child's teacher, when consulted, was able to solve the problem quite easily. The girl had undergone, at the beginning of the year, an operation that left unsightly scars on her legs and rendered her temporarily incapable of taking a full part in the life of the school. She was a highly sensitive child and painfully conscious of her disability. Towards the end of the year she recovered to the extent that she was able to secure a place in the school's netball team, and the consequent increase in her self-confidence manifested itself in a marked improvement in all branches of her work in school, and so in her performance in the two intelligence tests given in the summer term. Clearly the impression formed by a panel of this girl's capabilities, if it were based on her test scores alone, would not be an accurate one; the evidence supplied by her teacher is indispensable if a true appraisal is to be made.

I am not, of course, suggesting that this method could or should be used for the assessment of every candidate, but its application to the group of children who constitute a border zone of the examination list could, I feel sure, help to reduce the number of wrongly allocated children.

SPECIAL ABILITIES

I have so far discussed the measurement of intelligence and various techniques for assessing attainment in English and Arithmetic. There are, of course, other abilities, discernible in children, that are usually grouped together and referred to as "special abilities" in order to distinguish them from the general ability that enters into all intellectual activities, so making a considerable contribution to success in the basic subjects of the junior school curriculum.

In failing as yet to give any consideration to the nature of the special abilities and to the means that can be employed for assessing them, I am not likely to have occasioned any surprise. In any discussion about the measurement of children's abilities, attention is paid almost exclusively to tests of intelligence and of attainment in English and Arithmetic. The reason for this is not far to seek. Many people associate the assessment of children's abilities almost exclusively with the examinations held by local education authorities for purposes of allocation to secondary school courses. This is regarded as *the* occasion for the measurement of a child's capacities and achievements. Since this examination rarely takes cognizance of special abilities, they tend to escape attention.

The emphasis placed on the assessment of general

ability in the examination of children of eleven can be attributed to two causes. The first is that at the junior school stage this ability plays a much greater part in all scholastic activities than do the special abilities, some of which, at the age of ten or eleven are only just emerging. A second and potent reason is that the secondary school entrance examinations still tend to be regarded as a device for separating eleven-year-old children into two groups: those capable of profiting from an academic course of education in a grammar school, and those unfitted for this type of education. Accordingly the form of examination used is one that has been found to achieve this desired dichotomy with a reasonable degree of accuracy.

Before discussing whether, even for this limited purpose, an assessment of special abilities should be disregarded, let us consider what some of these abilities are. There is some disagreement about the number of abilities that can be regarded as special rather than general and as definitely established. It is not always easy to draw, with confidence, a firm boundary between two groups of activities and regard them as relatively independent and isolable; they may seem to be so in one group of children, whereas in another they may be found to correlate highly with each other and appear therefore to be aspects or expressions of the same underlying ability. Few would deny, however, that the following distinguishable abilities can be detected in some children:

1. Verbal Ability

Children of equal intelligence, as measured by individual tests or by "non-verbal" tests, are often found to differ markedly in the size of their effective vocabulary, their ability to use words in speech or writing and, eventually, in the degree of success that they are capable of achieving in linguistic and literary studies.

2. Musical Ability

Some children are found to be more capable than others of fine discriminations in response to variations in the pitch and intensity of musical notes. Some develop a marked sense of rhythm and harmony and remember melodies easily. These differences justify the postulation of a specific musical ability that manifests itself in the flair that some enjoy for learning to play musical instruments or in later years for the capacity for critical appreciation.

3. Artistic or Aesthetic Ability

Irrespective of the amount of training received, some children are able to draw or paint more successfully than others and betray a richer appreciation of colour and form than their less gifted contemporaries.

4. Mechanical Ability

Just as certain gardeners are accused by envious neighbours of possessing "green fingers," so there

are people who appear to develop a flair for understanding and manipulating mechanisms of all kinds.

5. *Spatial Ability*

This is the ability to perceive and deal with spatial relationships and involves memory for shapes and the ability to turn them, in imagination, into different positions and to form a visual image of their appearance from any angle. This ability, in which there are wide individual differences, probably makes a contribution to performance in the exercise of some mechanical and artistic skills.

6. *Numerical Ability*

This is the specific ability that enters into all arithmetical processes and enables some children to be better "at figures" than others who may be their equals in general ability.

* * * *

There are tests available that measure, or purport to measure, with varying degrees of reliability, all the abilities listed above, and even for those abilities—particularly in music and art—for which the existing tests are unsatisfactory, some assessment can be made which will at least serve to distinguish the highly gifted from those of average or less than average endowment.

Is there any justification for taking into consideration estimates of any or all of these abilities at eleven or at some other stage in a child's educational career?

I would suggest two possible arguments in favour of making such an attempt.

In the 1944 Education Act, the provision is recommended of a variety of secondary school courses sufficient in number to cater for the discernible needs and aptitudes of all primary school leavers. The duty would therefore appear to be imposed upon us of exploring the whole range of special abilities in order to determine what courses are needed in order that adequate opportunities can be afforded to each child for the full development of his individual talents.

A second reason for taking special abilities into consideration seems to me to be valid even in the case of children who are to follow the same curriculum. Consider as an example a number of children who all fall within a narrow range of general ability and have gone, say, to a grammar school. Suppose that by means of additional tests we have discovered that half of these children excel in verbal ability, and that the other half, whilst not markedly gifted in this direction, have considerable spatial ability. Quite apart from the fact that this finding might justify introducing some change into the curriculum followed by the two groups—extra languages for the former, for example, and greater emphasis on some branches of mathematics and applied science for the latter—is it not also desirable that even the same subjects should be presented to the two groups in different ways? It might well be that in dealing with the same subject matter the spatially-gifted group

could reach the same level of comprehension by the study of models, charts, graphs and other "visual aids," as that reached by those children of greater verbal facility who are taught orally. Every good mathematics teacher acts on the principle in dealing with his "C" forms.

I need not say how important spatial ability is when it comes to diverting children into what are called "technical" courses rather than "academic" courses.

In discussing special abilities I have refrained from commenting on the relative contributions to their development made by heredity and experience. It is difficult to disentangle these influences and the task is further complicated by the fact that in two individuals the same degree of skill may be achieved as a result of different proportions of the two contributory factors. For practical purposes, however, precise knowledge of the ingredients and life history, so to speak, of a particular talent is of little importance. The important fact, that has to be taken into account, is that in some individual special abilities emerge and tend, often, to persist.

LOOKING TO THE FUTURE

This book had its origin in a conversation, in which a parent of a ten-year-old boy and the child's teacher questioned me about the methods that are commonly employed in assessing the abilities of primary school leavers in order to allocate them to suitable courses of secondary education.

I have tried to answer all the questions that they put to me. In doing so, I have outlined the principles governing the construction and standardisation of intelligence tests; I have discussed some of the factors, including coaching and practice, that can affect a child's performance in these tests; I have examined the methods currently used for measuring children's achievements in English and Arithmetic; I have considered the advantages and disadvantages of using an interview and of inviting teachers to furnish assessments of their pupils' abilities; and I have commented on some of the abilities and characteristics manifested by eleven-year-old children that are not at the moment being sufficiently taken into account in estimating their educational needs and potentialities.

I have discussed what is done, and some of the things that might be done. Now I should like to outline what I think *will* be done in the near future. In other words, I am not proposing to outline what is, necessarily, my notion of the ideal scheme for dealing with the problem of assigning children to suit-

able courses of secondary education, but rather to try to forecast the outcome of what I conceive to be the present trends.

First, let me describe a dream—or perhaps it would be more fittingly described as a nightmare.

I was sitting, in my dream, with about forty other patients in the large waiting-room of a doctor's surgery. We all knew the doctor well. He had attended us all for years, understood our symptoms, and had treated us successfully during many an illness. We were, nevertheless, anxious and tense on this particular occasion. For this was no ordinary day. This was the day of *The Diagnosis*. A bell rang, the subdued conversation died away and we filed into the consulting-room. The doctor was there, smiling upon us all and murmuring words of encouragement. The table in front of him was piled high with little sealed boxes containing the instruments sent by the Local Health Authority. On the outside of each box was a printed guarantee to the effect that each instrument was reliable and standardised.

We sat down, and the doctor unfastened the boxes, passed the thermometers round and looked at his stop-watch. This morning, as we all knew, we were to have the Temperature Test; in the afternoon we would all return for Blood Pressure and Pulse Rate. I remember feeling irritated as I glanced across at the smug, flushed face of my neighbour; he had been taking some pills that were reputed to raise the temperature to 102 within a period of three hours.

I remember little of the rest of the tests, but the anxious weeks of waiting for the results stand out quite vividly in my recollection. One day during this period I overheard the doctor talking to a colleague. I had never heard him sound quite so irritated before. "Twenty-five years," he was saying, "Twenty-five years I've been a general practitioner and they can't trust me to say whether a patient should go to a sanatorium, a mental hospital or a surgical ward."

At last the results came through. The doctor read them out to the assembled company. Seven of us were to go to the big new hospital (including my neighbour who had taken the pills). The rest were to be discharged as incurable.

This is the sort of dream that does not disturb one's waking thoughts. Clearly this dire situation could never come about. The medical profession is far too powerful to allow anything like this to happen.

It is, of course, with the parallel situation in our own profession that I am concerned. In many areas teachers are being asked to step aside when it becomes necessary to effect a diagnosis of their pupils' abilities, and only a few feeble protests have so far been heard either from teachers themselves or from the public. I feel convinced, however, that there is a growing tendency to question the wisdom of excluding the teachers from an active participation in the decisions that have to be made about the educational future of eleven-year-old children. I am equally convinced that the time is not far distant

when teachers will find themselves fully responsible for the allocation of their pupils to appropriate courses of secondary education, except that it may be found necessary to submit the case of an occasional "border line" candidate to a panel of referees.

Two changes, it seems to me, are required in order to facilitate this development. One can only be carried out by the local education authorities; the other is in the power of teachers themselves to effect.

Steps need to be taken by the education authorities to assuage parental anxiety about the transition from primary to secondary education. In far too many areas competition for a few coveted grammar school places is fierce and a child who fails to obtain one of them is often regarded as having been deprived of a proper opportunity to realise his talents.

The belief that the grammar school provides the only avenue to a satisfying career and the only environment in which a child can develop his intellectual powers to the full, needs not only to be attacked by argument but also to be disproved by demonstration. The increase in the number of grammar school places available in certain under-privileged areas until it approaches the national average, the provision of academic courses in some, at any rate, of the secondary modern schools for those children who can prove their capacity to profit from them, the endowment of *all* secondary schools with some articulate purpose and recognisable and worthy educational objectives—these are some of the measures that need to be taken in order that the

transfer of a child from a primary school to a secondary school may become a smooth transition rather than a period of crisis.

The apparent finality of the decisions made about the appropriate educational treatment of eleven-year-old children is another source of anxiety to parents. A flexible system of transfer between various types of secondary school course should not only be devised but should be operated and be seen to operate satisfactorily.

The second change to which I referred, and which is one that only teachers themselves can introduce, concerns the status and rôle of teachers as conceived by the public in general and by parents in particular. The doctor's proffered diagnosis of his patient's condition is but rarely challenged. From time to time mistakes are made, and criticisms levelled against them but, for the most part, doctors make strong claims on the public's respect and regard. They are recognised as skilled, adequately trained individuals, who practise their art, moreover, in accordance with the firm dictates of a strict code of professional ethics. It is scarcely conceivable that they could be suspected of allowing their clinical judgments to be distorted in order to please one person or to spite another.

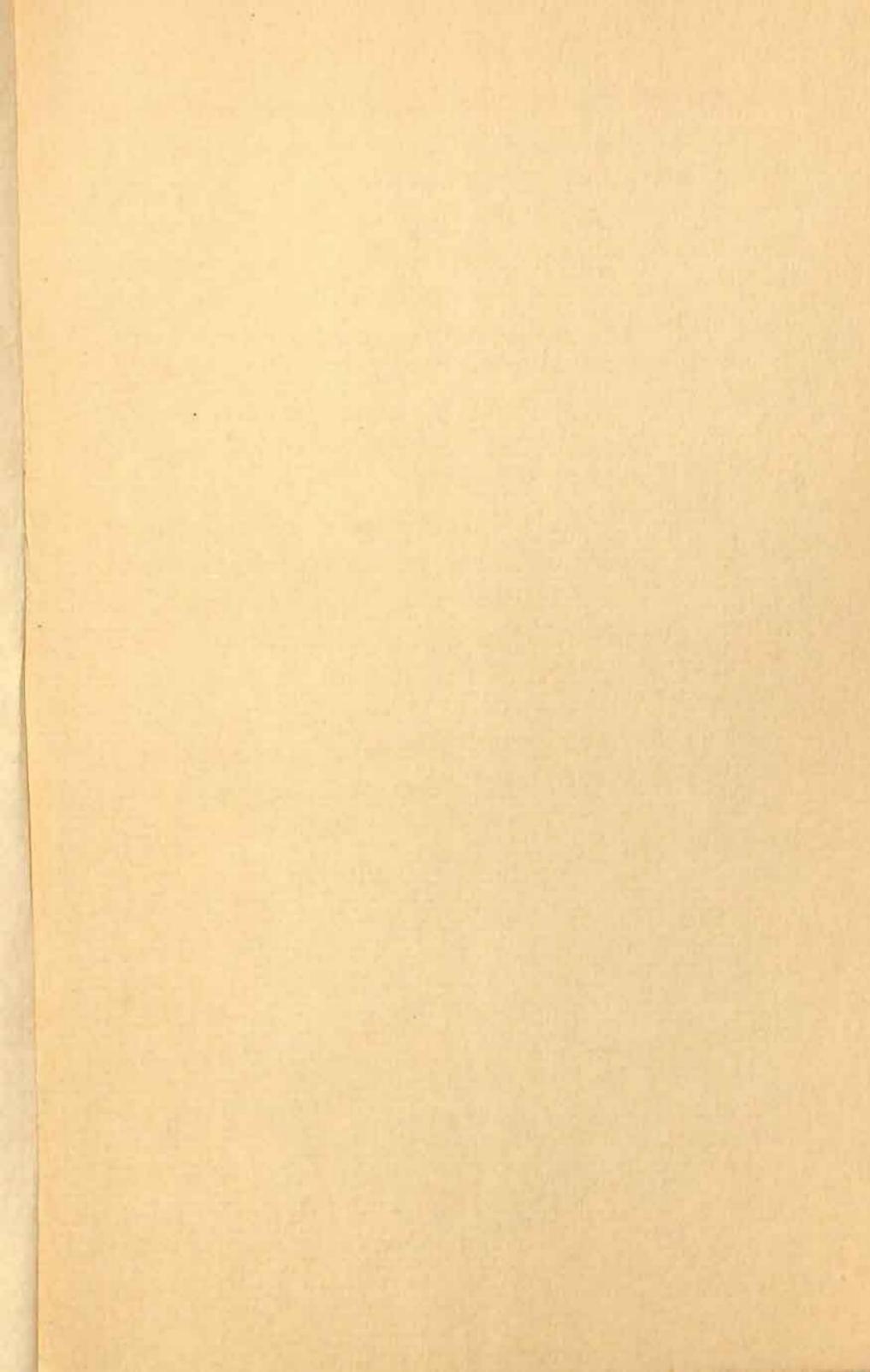
Teachers do not as yet enjoy an equivalent measure of public esteem. This cannot be won by persuasion or argument. It must be earned. If teachers are to prove equal and be judged equal to the responsibility that I feel sure is soon to be theirs,

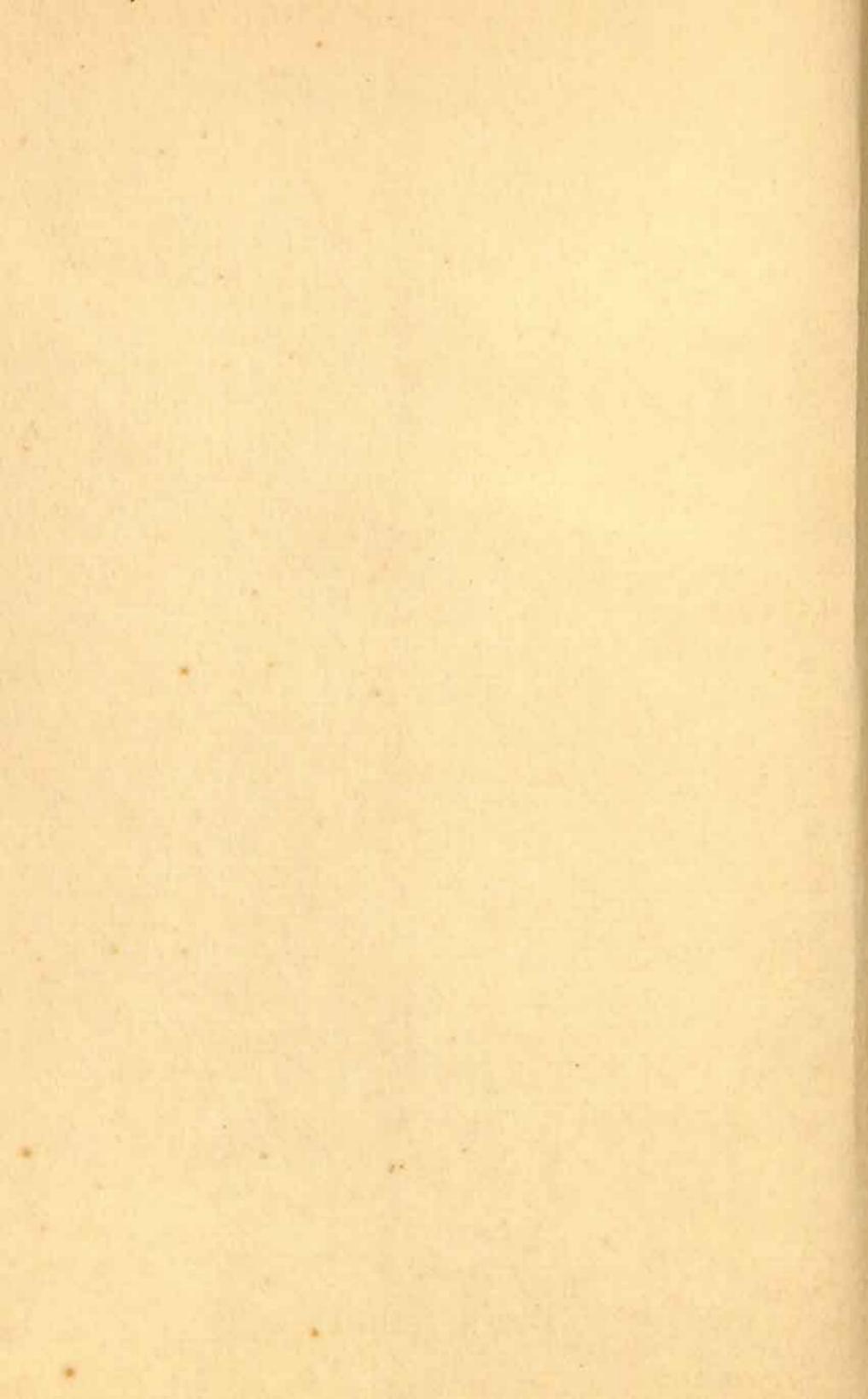
they must so prepare and equip themselves as to show that they are capable of assessing a child's abilities and aptitudes with impressive accuracy, and of prescribing for him a form of educational treatment that is demonstrably appropriate. To do this they must become as familiar with the available diagnostic tools as the general practitioner is with his thermometer and stethoscope, and be able to interpret their pupils' records with the same degree of clinical flair that the doctor applies to his patients' case histories.

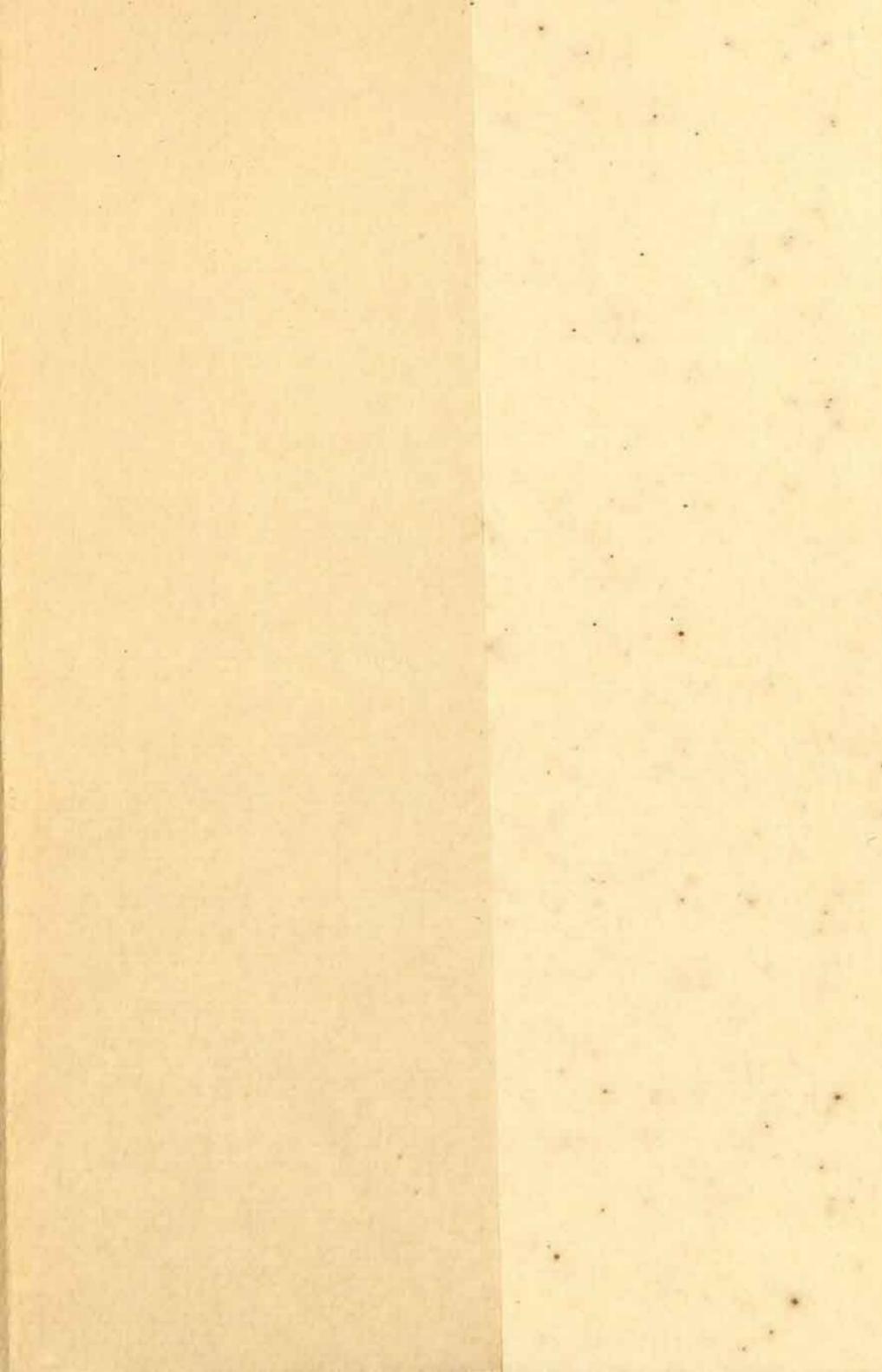
I believe that these changes that I have outlined can and will take place. Public reaction to them will depend upon the skill, courage and integrity with which teachers perform the task for which their professional training and experience should have fitted them.

I think that there is every hope of substantial improvements in what we are doing for the children of this country, upon whom a solution to the difficult tasks ahead of us as a nation must wholly depend.









"Are you able to assure me that the sort of examination which my child will be taking shortly will reveal his abilities in their true form?"

This question, asked by so many parents, has been the source of endless controversy between teachers, educationalists, and psychologists ever since it was first claimed that the abilities of children could be measured with something like accuracy.

Dr. Watts, whose wide experience in education as teacher, university lecturer, inspector of schools, and educational research worker, makes him especially fitted to answer this question, has tackled the problem in a straightforward and honest manner. His clear and interesting discussion of this vital subject will help those who read this book to obtain a better understanding of the aims, methods, and actual techniques of mental testing.

UNIVERSITY OF LONDON PRESS LTD